

Implementation of Backpropagation Neural Network and Blood Cells Imagery Extraction for Acute Leukemia Classification

¹Faisal Asadi, Fiqhri Mulianda Putra, Mutik Indah Sakinatunnisa, Fadhilah Syafria, and Okfalisa

¹Department of Informatics Engineering
State Islamic University of Sultan Syarif Kasim Riau
Pekanbaru, Indonesia

^{1,2}Ismail Marzuki

Department of Computer Engineering
Kwangwoon University
Seoul, South Korea

Abstract— This paper proposes an implementation of classification of Acute Leukemia using backpropagation neural network algorithm and blood cells imagery extraction. Leukemia is a cancer of blood cells, known as an abnormal white blood cells growth produced by the bone marrow. The exact cause of leukemia is still unknown. However, leukemias have been acknowledged to be grouped by how quickly the disease develops (acute leukemia and chronic leukemia) as well as by the type of blood cell that is affected (lymphocytes or myelocytes). This paper focuses on the acute leukemia, which can be categorized into Acute Lymphoblastic Leukemia (ALL) and Acute Myelogenous Leukemia (AML). These types of leukemia are possible to be diagnosed by counting the number of blood cells growth in the bone marrow through the microscopic analysis of blood cell imagery. However, it cost overpriced in terms of time, energy, and amount. In addition, manual counting may lead potential false of diagnoses. In this paper, backpropagation neural network algorithm is used to extract the characteristics of ALL and AML blood cells. Digital image processing is employed for identification type of leukemias. The experimental results argue that the proposed work achieves about 86.66% accuracy on average in classifying the leukemia acute types.

Keywords— Acute Leukemia Classification; Backpropagation Neural Network; Blood Cells Extraction; Digital Image Processing

I. INTRODUCTION

Leukemia is a blood cell cancer, specifically it is a cancer of the white blood cells. This cancer is recognized as the number of white blood cells grow abnormally in the bone marrow. Therefore, it is also sometime called as a bone marrow cancer. The bone marrow is the flexible tissue in the interior of human bones which produces three types of blood cells, such as white blood cells (leukocyte), red blood cells (erythrocyte), and platelets (thrombocyte). For human body, leucocyte is known as the cells of the immune system involved in protecting the body against both infectious disease and foreign invaders. Erythrocyte is used for delivering oxygen to the body tissues via blood flow through the circulatory system. Thrombocyte is to stop bleeding by clumping and clotting blood vessel injuries.

Leukemia is often described as being either acute (fast growing) and chronic (slow growing). Based on its cell morphology, leukemia can be differed as Acute Leukemia

Lymphoblastic (ALL) and Acute Myeloid Leukemia (AML). Different types of leukemia have different treatment options and outlooks. Generally, it will need long time to conduct leukemia diagnoses so-called gold standard procedures. It starts by checking the bone marrow aspiration and capturing the chest radiography. In addition, cerebrospinal liquid and other medical check-ups might also necessary. This gold standard procedure can diagnose about 90% case of leukemias. The remaining cases will need additional procedures such as cytochemical, immunology, cytogenetic, and molecular biology [1, 2]. In Indonesia, the Ministry of Health reports that leukemia is the biggest number of blood cancer in Indonesia among other blood cancer types; And majority this cancer is suffered by children. In 2015, there were about 77% ALL patients in age 2-6-year-old, 11% patients suffered AML, and the rest of patients had Chronic Myeloid Leukemia [3].

There have been many studies researching to find the best way in diagnosing leukemia. Adnan, et al [4] has studied to identify ALL by using blood smear images and a neural classifier. This research employed classification technique of the backpropagation method with 900 layers and 62 hidden layers. This study claimed about 96% satisfaction in identifying the ALL cases. Subrajeet, et al [5] has published similar study by combining Functional Link Artificial Neural Network (FLANN) and imagery extraction technique. The study aimed to segment Lymphoblast images based on the white blood cell imagery. In our proposed work, it is proposed to classify and identify Acute Leukemia. We use backpropagation neural network algorithm to extract the characteristics of blood cells ALL and AML, and digital image processing for the leukemia identification technique. The experimental results show about 86.66% accuracy in classifying the leukemia acute types.

The rest of this paper is organized as follows. Section 2 presents the overview of backpropagation and digital image processing used for this proposed work. In Section 3, the proposed work is presented. Section 4 shows the performance evaluation of the proposed work. And finally, Section 5 concludes this work with further research topics.

II. OVERVIEW OF BACKPROPAGATION AND BLOOD CELLS IMAGERY EXTRACTION FOR ACUTE LEUKEMIA CLASSIFICATION

A. Backpropagation Neural Network

Backpropagation is one of artificial neural network models. It is usually used to train a neural network model to get a balance condition of a network capability in recognizing patterns for training stage and in providing correct response to input data of a network model for testing data stage. Input data model in both training and testing stage are usually similar [6]. In general, backpropagation utilizes gradient descent algorithm. It is known as a Multi-Layer Network. It is due to the backpropagation is formed by generalizing the training rules in model Widrow-Hoff. This algorithm creates so-called hidden layer. Accordingly, the backpropagation architecture is recognized to have input, hidden, and output layer. Every neuron in input layer is connected to each neuron in hidden layer. And neuron in output layer is linked with all neurons in hidden layers. Figure 1 depicts an illustration of this backpropagation neural network architecture.

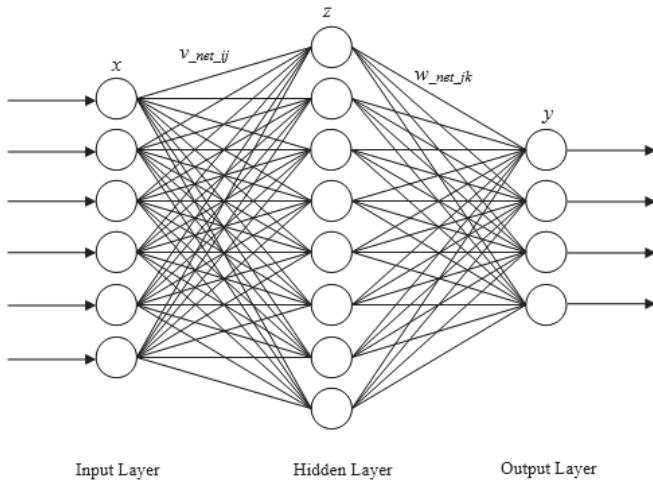


Fig. 1. Backpropagation neural network architecture

B. Hue Saturation Value (HSV)

HSV color space model is known as a color representation which defines color based on terminology of Hue, Saturation, and Value. Hue is used to distinguish colors representing redness and greenness of lights and so forth. Saturation denotes a pureness level of color which indicates how much white color is given to a color. Value is an attribute of lights which can be received by human eyes regardless what the color is. HSV color space has been widely used in digital image processing in order to covert Red, Green, and Blue color space (RGB) to HSV. HSV before the conversion to RGB can be expressed as follows:

$$H = \tan \left[\frac{3(G - B)}{(G - B) + (R - B)} \right] \quad (1)$$

$$S = 1 - \frac{\min(R, G, B)}{V} \quad (2)$$

$$V = \frac{R + G + B}{3} \quad (3)$$

To convert RGB to HSV color space, the value of red, green, and blue need to be priority normalized which can be defined as:

$$r = \frac{R}{R + G + B} \quad (4)$$

$$g = \frac{G}{R + G + B} \quad (5)$$

$$b = \frac{B}{R + G + B} \quad (6)$$

where R, G, B denote red, green, and blue color of a pixel before normalization; and r, g, and b express the normalized value of RGB, respectively. Therefore, the conversion formula of RGB color space to HSV color space is formed as:

$$V = \max(r, g, b) \quad (7)$$

$$S = \begin{cases} 0 - \frac{\min(R, G, B)}{v}, & \text{if } V = 0 \\ V - \frac{\min(R, G, B)}{v}, & \text{if } V > 0 \end{cases} \quad (8)$$

$$H = \begin{cases} 0, & \text{if } S = 0 \\ 60 \times \left[0 + \frac{g - b}{S \times V} \right], & \text{if } V = r \\ 60 \times \left[2 + \frac{b - r}{S \times V} \right], & \text{if } V = g \\ 60 \times \left[4 + \frac{r - g}{S \times V} \right], & \text{if } V = b \\ H + 360, & \text{if } H < 0 \end{cases} \quad (9)$$

The converted RGB values to HSV can be usually used for a color feature extraction in digital image processing for further analysis depend on what application is being developed. For this purpose, mean formula is employed to perform the color feature extraction by using HSV color space. The mean equation can be found as:

$$\mu = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N p_{ij} \quad (10)$$

C. Gray Level Co-Occurrence Matrix (GLCM)

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. GLCM can be computed to show relationship of two adjacent pixels with a certain intensity, distance, and angle. Distance in pixel-based is usually expressed as 1, 2, 3, and so forth. Angle orientation defines a degree value which can be 0°, 45°, 90°, or 135°. This value represents the gray-level value transformed into a co-occurrence matrix. This transformation is obtained with a sliding window which can be in size 3×3, 5×5, 7×7, 9×9, and so on. This window can be then to determine a spatial relationship of distance and angle. Figure 2 illustrates a 3×3 window size where showing the corresponding of distance and angle (bv) of this matrix.

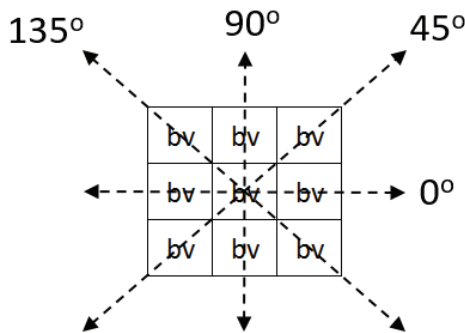


Fig. 2. Corresponding of distance and angel of 3×3 matrix

To get the gray-level co-occurrence matrix, an image with RGB color space needs to be converted first to grayscale image. This conversion can use the following formula:

$$Y_{rgb} = 0.299 \times R + 0.587 \times G + 0.114 \times B \quad (11)$$

The following statistic equations are used in this study to get the texture feature with GLCM:

- Angular Second Moment (ASM)

$$ASM = \sum_i \sum_j \{p(i,j)\}^2 \quad (12)$$

- Contrast

$$CON = \sum_k k^2 \left[\sum_i \sum_j p(i,j) \right]_{|i-j|=k} \quad (13)$$

- Correlation

$$COR = \frac{\sum_i \sum_j (i,j) \times p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (14)$$

- Variance

$$VAR = \sum_i \sum_j (i - \mu_x) \times (j - \mu_y) \times p(i,j) \quad (15)$$

- Inverse Different Moment

$$IDM = \sum_i \sum_j \frac{1}{1 + (i - j)^2} \times p(i,j) \quad (16)$$

- Entropy

$$ENT = - \sum_{i,j} p(i,j) \times \log p(i,j) \quad (17)$$

D. Acute Lymphoblastic Leukemia (ALL)

ALL is one of leukemia diseases which forms normal blood cells into a deadly lymphoblastic. This lymphoblastic will then quickly grow its number of the blood cells in the bone marrow. The following figure shows how the acute lymphoblastic leukemia grows in the human's bone marrow [1].

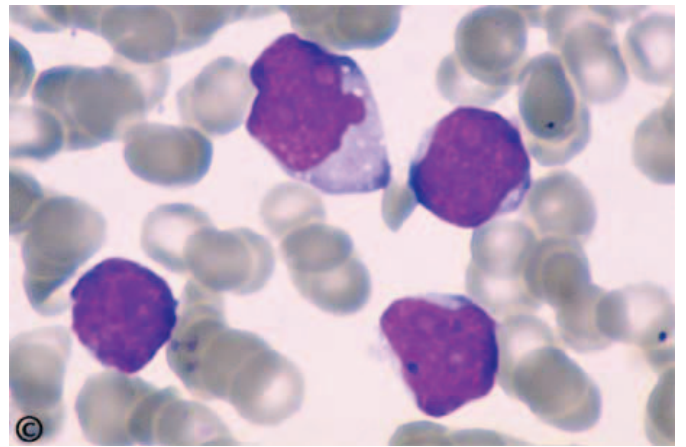


Fig. 3. Blood cells of Acute Lymphoblastic Leukemia (ALL)

E. Acute Myelogenous Leukemia (AML)

AML is a blood cancer which is formed similar with myeloid of white blood cells. It can be identified based on its breeding as an abnormal blood cells in the bone marrow. However, it is generally formed and grown together with the white blood cells. AML usually strikes adults and is more often found in men bodies than women. Figure 4 shows the AML in bone marrow of human [1].

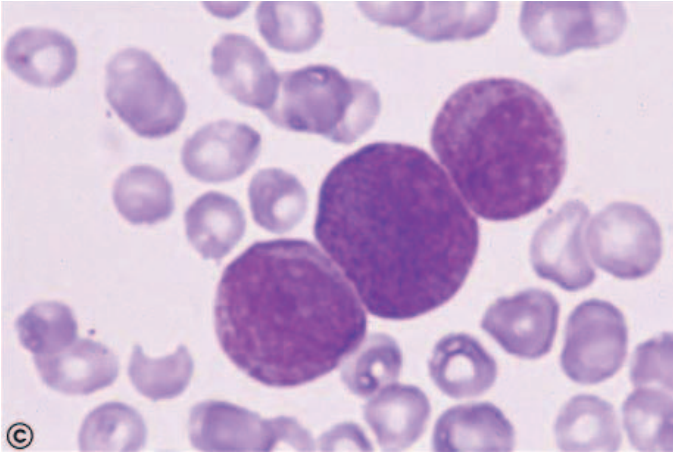


Fig. 4. Blood cells of Acute Myelogenous Leukemia (AML)

III. THE PROPOSED WORK

In this section, our implementation of backpropagation neural network algorithm and blood cells imagery extraction technique for classification of acute leukemia is discussed as follows:

- 1) *Data collecting.* We firstly collect the blood cells image as sample data for both training and testing data from [7]. There are 75 samples of blood cells images captured from human's bone marrow. These images have been scaled by using Zeiss microscopic and Nikon with scale size: $\times 200$, $\times 400$, $\times 630$, $\times 1000$.
- 2) *Preprocessing.* Preprocessing techniques used in this work are cropping and resizing to all the sample data. This preprocessing output is only focused on one single cell with size 300×300 pixel. However, it should be noted that we do not do any background removal technique into our sample data and the cytoplasm is also kept from the entire images of blood cells.
- 3) *Digital image analysis.* In this step, we converted the RGB images to HSV color space by using (4) to (9).
- 4) *Feature extractions.* As mentioned before, there are two kinds of feature extraction used in this study: color feature extraction and texture feature extraction. For color extraction, HSV color feature is used and for texture feature extraction, we employ GLCM. Both techniques are performed onto all sample data. In the color extraction, we used (10) as further analysis after computing (4) to (9) in the third stage. In the texture extraction, we computed all formulas from (11) to (17) in this stage. Figure 5 and Figure 6 are given to show how both feature extractions were performed in this proposed work.
- 5) *Training and testing of classification.* In this stage, we do the training and testing stages to classify the acute lymphoblastic leukemia and the acute myelogenous leukemia by using backpropagation neural network. Training stage is done by input each leukemia cell images. Then, several parameters are adjusted, including

learning rate, epoch, minimum target error, and number of neuron. The best parameters are then used for testing stage. We used our data in composition as 80% for training data and 20% as testing data or 60 training data and 15 testing data.

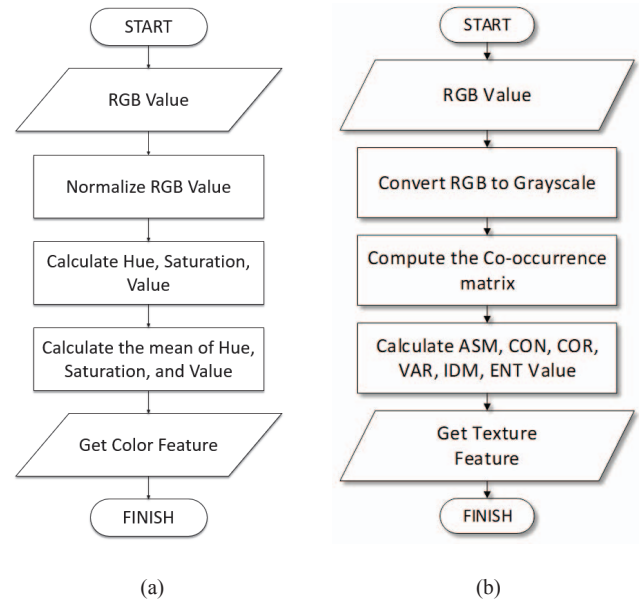


Fig. 5. Flowchart of (a) HSV color feature extraction and (b) texture feature extraction by using GLCM

IV. IMPLEMENTATION AND EXPERIMENT RESULTS

In this section, we discuss the results of our implementation of backpropagation neural network algorithm and blood cells imagery extraction technique for classification of acute leukemia. Figure 7 and Figure 8 show the user interface of neural network training and application of this implementation, respectively. In this study, we use MATLAB to develop our proposed work. To evaluate our proposed work, we calculate the accuracy (Acc) of our model by using formula defined as:

$$Acc = \frac{\sum \text{True positive value}}{\sum \text{Total Data Test}} \times 100\% \quad (18)$$

After we trained our data sample, we found the best parameter for our model, then we used the parameters for data testing that can be detailed as follows: Epoch (the number of iteration) is set to 5, Minimum error is set to 0.001, Learning rate equals to 0.01, and input layer is 10.

The results of our implementation are tabulated as in Table 1 and Table 2. Table 1 shows the classification testing of ALL and Table 2 presents the classification testing of AML, respectively. According Table 1, there are 14 cases of 15 identified as ALL by using our approaches. In this study, ALL can be detected with satisfaction score about 93.33%. However, when tested our algorithm to detect AML cases, we only achieved about 80% satisfaction or 12 cases are true out of 15 cases. Consequently, we can report that on average, our

algorithm can argue about 86.66% accuracy in classifying and identifying the acute leukemia cases.

V. CONCLUSION

In this paper, an implementation of backpropagation neural network and blood cells extraction imagery to classify acute leukemia is proposed. We combine these two approaches and test our algorithm by using 75 data samples. In our implementation, we used backpropagation neural network algorithm to extract the characteristics of blood cells ALL and AML and digital image processing for the leukemia identification technique. The experimental results show that our algorithm can identify ALL and AML cases about 93.3% and 80%, respectively, or we can achieve about 86.66% accuracy in classifying the leukemia acute types on average.

REFERENCES

- [1] T. Haferlach, W. Kern, S. Schnittger, and C. Schoch, "Modern diagnostics in acute leukemias," *Journal of Critical Reviews in Oncology/Hematology*, vol. 56, no. 02, pp. 223-234, October 2005.
- [2] G. J. L. Kaspers and C. M. Zwaan, "Pediatric acute myeloid leukemia: towards high-quality cure of all patients," *Hematological: Journal of the European Hematology Association*, vol. 92, pp. 1519-1532, November 2007.
- [3] Jurnal Infodatin kanker, Pusat Data dan Informasi Kesehatan Kementerian Kesehatan Republik Indonesia, ISSN 2088-270X Semester 1, 2015.
- [4] A. Khasman and H. H. Abbas, "Acute lymphoblastic leukemia identification using blood smear images and a neural classifier," *International Work-Conference on Artificial Neural Networks*, pp. 80-87, 2013.
- [5] S. Mohapatra, D. Patra, S. Kumar and S. Satpathy, "Lymphocyte image segmentation using functional link neural architecture for acute leukemia detection," *Biomedical Engineering Letters*, vol. 02, pp. 100-110, 2012.
- [6] S. Mohapatra, P. K. Sa, and B. Majhi, "Adaptive threshold selection for impulsive noise detection in images using coefficient of variance," *Neural Computing and Applications*, vol. 21, no. 02, pp. 281-288, March 2012.
- [7] N. Medeiros, <http://hematologyatlas.com/principalpage.htm>

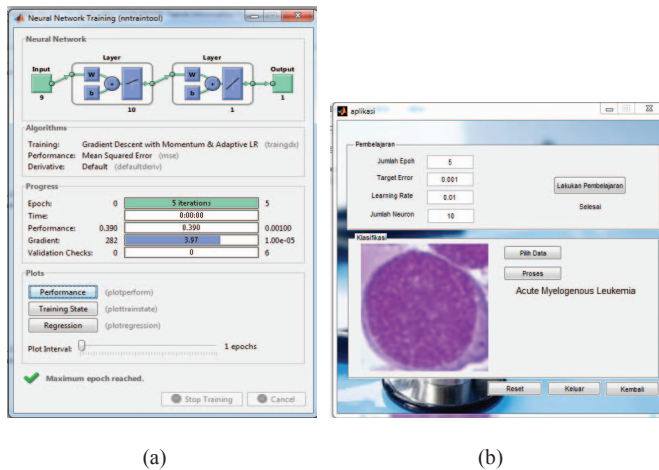


Fig. 6. User interface of (a) neural network training and (b) acute leukemia classification

TABLE I. TESTING OF ACUTE LYMPHOBLASTIC LEUKEMIA (ALL)

Data testing	Detected	Not Detected	Execution Time (second)
ALL-1	detected	-	0.557280
ALL-2	detected	-	0.109295
ALL-3	detected	-	0.114100
ALL-4	detected	-	0.110911
ALL-5	detected	-	0.114721
ALL-6	detected	-	0.556588
ALL-7	detected	-	0.113882
ALL-8	-	not detected	0.102314
ALL-9	detected	-	0.106794
ALL-10	detected	-	0.109938
ALL-11	detected	-	0.106026
ALL-12	detected	-	0.105420
ALL-13	detected	-	0.107634
ALL-14	detected	-	0.11570
ALL-15	detected	-	0.115831

TABLE II. TESTING OF ACUTE MYELOGENOUS LEUKEMIA (AML)

Data Testing	Detected	Not Detected	Execution Time (second)
AML-1	detected	-	0.558305
AML-2	detected	-	0.557827
AML-3	detected	-	0.109747
AML-4	detected	-	0.101995
AML-5	detected	-	0.105071
AML-6	detected	-	0.127170
AML-7	detected	-	0.097921
AML-8	-	not detected	0.106153
AML-9	detected	-	0.123054
AML-10	detected	-	0.105012
AML-11	-	not detected	0.118990
AML-12	-	not detected	0.118436
AML-13	detected	-	0.106185
AML-14	detected	-	0.104136
AML-15	detected	-	0.103803